Pair Correlation Function Integrals

Computation and Use

Wedberg, Nils Hejle Rasmus Ingemar; O'Connell, John P.; Peters, Günther H.J.; Abildskov, Jens

Published in:
Journal of Chemical Physics

Link to article, DOI:
10.1063/1.3626799

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Pair correlation function integrals: Computation and use

Rasmus Wedberg,[1] John P. O'Connell,[2] Günther H. Peters,[3] and Jens Abildskov[1,a)]
[1]*CAPEC – Department of Chemical and Biochemical Engineering, Søltofts Plads, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[2]*Department of Chemical Engineering, University of Virginia,102 Engineer's Way, PO Box 400741, Charlottesville, Virginia 22904-4741, USA*
[3]*MEMPHYS – Center for Biomembrane Physics, Department of Chemistry, Kemitorvet, Building 207, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

(Received 20 April 2011; accepted 29 July 2011; published online 24 August 2011)

We describe a method for extending radial distribution functions obtained from molecular simulations of pure and mixed molecular fluids to arbitrary distances. The method allows total correlation function integrals to be reliably calculated from simulations of relatively small systems. The long-distance behavior of radial distribution functions is determined by requiring that the corresponding direct correlation functions follow certain approximations at long distances. We have briefly described the method and tested its performance in previous communications [R. Wedberg, J. P. O'Connell, G. H. Peters, and J. Abildskov, Mol. Simul. **36**, 1243 (2010); Fluid Phase Equilib. **302**, 32 (2011)], but describe here its theoretical basis more thoroughly and derive long-distance approximations for the direct correlation functions. We describe the numerical implementation of the method in detail, and report numerical tests complementing previous results. Pure molecular fluids are here studied in the isothermal-isobaric ensemble with isothermal compressibilities evaluated from the total correlation function integrals and compared with values derived from volume fluctuations. For systems where the radial distribution function has structure beyond the sampling limit imposed by the system size, the integration is more reliable, and usually more accurate, than simple integral truncation. © *2011 American Institute of Physics*. [doi:10.1063/1.3626799]

I. INTRODUCTION

Integrals of the total correlation function provide a direct link between the microscopic structure of a fluid mixture and macroscopic thermodynamic derivative properties by a set of algebraic relations termed fluctuation solution theory,[1–3] or Kirkwood-Buff theory. Isothermal compressibilities, partial molar volumes, and composition derivatives of activity coefficients are expressed in terms of these integrals.[4] The integrals have been extensively used in the development of accurate force fields for fluid mixtures.[5–12] Christensen *et al.*[4,13,14] utilized the integrals in a methodology for predicting the excess Gibbs energy of liquid mixtures. They simulated the mixture at a few compositions using molecular dynamics and then regressed parameters of the modified Margules G^E model.[15] This is an alternative approach to free energy calculations such as Widom insertion,[16] thermodynamic integration,[17–19] and Gibbs-ensemble Monte Carlo.[20–22] The method furthermore suppresses statistical errors in the functions since the derivatives are used to obtain the parameters. For substances exhibiting nearly ideal behavior in the vapor phase, this approach has proven to be a viable alternative for predicting vapor-liquid equilibria.[13,14]

In addition, correlation function integrals themselves are interesting properties as several established corresponding state theories for thermodynamic properties are based on modeling direct correlation function integrals.[23–29]

Numerical integration of the radial distribution functions from molecular simulation is often inaccurate because the integrals rarely converge adequately for the system sizes commonly used. Partly this is due to the functions retaining subtle structure over relatively long distances that contributes to the integral. A technique due to Theodorou and Suter[30] allows calculation of the pair correlation functions for distances up to $\sqrt{3}/2$ times the box dimension. This might however still be insufficient for convergence, and the pair correlation functions obtained in this way often exhibit substantial noise as the upper limit is approached.[30] The role of finite-size effects on the accuracy of the functions as the sampling limit is approached has also been discussed.[31] Several approaches have been proposed in order to correct the sampled distribution functions and extend them to long range, so that the integrals can be evaluated accurately. Weerasinghe and Smith[5] evaluated the integrals numerically using an appropriately selected truncation radius. Perera and Sokolić,[32] and Hess and van der Vegt[33] corrected the sampled distribution functions by rescaling them such that they converged to unity within the sampling limit. Matteoli and Mansoori,[34] Christensen *et al.*,[4] and Wedberg, Peters, and Abildskov[35] fitted the functions to parametric expressions selected by empirical means, which then were used for extrapolation to long range. Another approach proposed by Nichols, Moore, and Wheeler[36] extrapolates the structure factors to zero with polynomials of varying degrees, in order to obtain the correlation function integrals. Approaches for correcting correlation functions specifically for the use of truncated potentials have also been discussed.[37,38]

a)Author to whom correspondence should be addressed. Electronic mail: ja@kt.dtu.dk.

A different approach to utilize density fluctuation expressions and correct for finite-size effects was recently described by Schnell *et al.*,[39] who calculated fluctuations in a sub-region of the simulation box. The size of this sub-region was extrapolated to infinity to obtain properties in the thermodynamic limit.

Our previous studies[40,41] extend molecular radial distribution functions by enforcing the corresponding direct correlation functions to follow a certain long-range approximation. This procedure is based on the work of Verlet[42] who applied the methodology to analyze simulations of the Lennard-Jones fluid. We applied the modified method to simulations of pure Lennard-Jones and Stockmayer fluids,[40] Lennard-Jones/Stockmayer mixtures,[41] and mixtures of water and alcohols,[41] demonstrating that integration of these extended functions yielded accurate derivative properties.

The present paper describes the basis more thoroughly including derivation of the direct correlation function approximations and the numerical implementation. Numerical tests comprise molecular dynamics simulations of the pure molecular fluids ethane, *n*-butane, *n*-hexane, 2-propanol and water, using the CHARMM force field,[43] in the isothermal-isobaric ensemble. The accuracy of the extended distribution functions is assessed by comparing isothermal compressibilities from the integrals with those from fluctuations in the simulation box volume. Comparisons are also made with integrals from the simple truncation approach by Weerasinghe and Smith.[5]

II. MOLECULAR CORRELATION FUNCTIONS

Unlike atomic total correlation functions, $h_{ij}(r)$, and pair radial distribution functions, $g_{ij}(r)$, which are functions of only the spatial distance r between the centers of mass of two molecules, molecular correlation functions depend on orientations, ω_1 and ω_2. As described in detail by Gray and Gubbins,[44] the total correlation function can be split into isotropic and anisotropic parts,

$$h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) = h_{ij}(r_{12}) + h_{ij}^{(\mathrm{a})}(\boldsymbol{r}_{12}\omega_1\omega_2). \tag{1}$$

The isotropic part $h_{ij}(r_{12})$ is identical to the atomic total correlation function, and is obtained by averaging out the angular dependence,

$$h_{ij}(r_{12}) \equiv \langle h_{ij}(r_{12}\omega_1\omega_2)\rangle_{\omega_1\omega_2}, \tag{2}$$

where

$$\langle\cdot\rangle_{\omega_1} \equiv \frac{1}{8\pi^2}\int d\omega_1 \equiv \frac{1}{8\pi^2}\int_0^{2\pi} d\phi_1 \int_{-1}^{1} d(\cos\theta_1)\int_0^{2\pi} d\chi_1, \tag{3}$$

and

$$\left\langle h_{ij}^{(\mathrm{a})}(\boldsymbol{r}_{12}\omega_1\omega_2)\right\rangle_{\omega_1\omega_2} = 0. \tag{4}$$

For flexible molecules, the correlation functions are also functions of the molecular conformations, but since the organic molecules considered in this work are small and do not (except *n*-hexane) possess multiple conformations, we ignore this effect.

The molecular Ornstein-Zernike equation defines the molecular direct correlation function, $c_{ij}(r_{12}\omega_1\omega_2)$,[44]

$$h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) = c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) + \rho\sum_l x_l \int \langle h_{il}(\boldsymbol{r}_{13}\omega_1\omega_3)c_{lj}(\boldsymbol{r}_{32}\omega_3\omega_2)\rangle_{\omega_3} d\boldsymbol{r}_3, \tag{5}$$

where ρ denotes the overall number density of the fluid and x_l is the number fraction of component l. In analogy with Eqs. (1), (2), and (4) the direct correlation function can be written as a sum of isotropic and anisotropic parts,

$$c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) = c_{ij}(r_{12}) + c_{ij}^{(\mathrm{a})}(\boldsymbol{r}_{12}\omega_1\omega_2). \tag{6}$$

Substituting Eqs. (1) and (6) into Eq. (5) and averaging out the angular dependence leads to

$$h_{ij}(r) = c_{ij}(r) + \rho\sum_l x_l \int h_{il}(r_{13})c_{lj}(r_{32})d\boldsymbol{r}_3 + \rho\sum_l \int \left\langle \left\langle h_{il}^{(\mathrm{a})}(\boldsymbol{r}_{13}\omega_1\omega_3)\right\rangle_{\omega_1} \left\langle c_{lj}^{(\mathrm{a})}(\boldsymbol{r}_{32}\omega_3\omega_2)\right\rangle_{\omega_2}\right\rangle_{\omega_3} d\boldsymbol{r}_3. \tag{7}$$

Neglecting the last term in Eq. (7) gives a simplified version of the Ornstein-Zernike equation, in which the isotropic direct and total correlation functions are related without the anisotropic terms,

$$h_{ij}(r) = c_{ij}(r) + \rho\sum_l x_l \int h_{il}(|\boldsymbol{r}-\boldsymbol{r}'|)c_{lj}(r')d\boldsymbol{r}'. \tag{8}$$

Here we use Eq. (8) instead of Eq. (5). While there may not be rigorous arguments for this choice, it is supported by several results. First, Eq. (8) is exact in several integral equation theories of fluids with anisotropic interactions, such as the mean-spherical approximation and the generalized mean field theory.[44] Second, Wang *et al.*[45] showed with Monte Carlo simulations of Lennard-Jones particles with significant dipole and quadrupole moments, that anisotropic forces have limited effects on $h_{ij}(r)$. Also, Gubbins and O'Connell[23] showed that for dense fluids the correlation function integrals for water and argon scaled with only two parameters. In addition, several studies show successful corresponding-states scaling for the direct correlation function integrals,[23–29] as described in detail in Ref. 46. Finally, the approximation of Eq. (8) is the first term of the spherical harmonic expansions of the molecular correlation functions.

III. METHOD

A. Extrapolation of molecular correlation functions

A method due to Verlet[42] aims to correct the correlation functions obtained from simulation for the effects of finite-sized systems, such as those summarized by Salacuse, Denton, and Egelstaff,[31] and to extend the correlation functions to longer ranges. The method was originally used for a

pure Lennard-Jones fluid, with focus on the qualitative behavior of the direct correlation function and the structure factor, and later extended to a Lennard-Jones mixture.[47] The present formulation of the method numerically determines total and direct correlation functions that satisfy the Ornstein-Zernike equation (8), under the constraints

$$\begin{cases} h_{ij}(r) = h_{\mathrm{MD},ij}(r), & r \le R_{ij} \\ c_{ij}(r) = t_{ij}(r), & r > R_{ij} \end{cases}, \tag{9}$$

where $h_{\mathrm{MD},ij}(r)$ are the total correlation functions obtained from simulation, $t_{ij}(r)$ are approximations of the long range part of the direct correlation functions and R_{ij} are the distances where the simulation correlation functions are matched to the calculated functions. Solving the Ornstein-Zernike equation with Eq. (9) yields correlation functions which may be integrated to obtain fluctuation properties. The procedure of solving the Ornstein-Zernike equation with the given constraints requires explicit approximations $t_{ij}(r)$ for the behavior of the direct correlation functions at large separations and choosing appropriate matching distances R_{ij}, as described later.

Commonly, the Wiener-Hopf factorization technique is applied when the direct correlation function is to be computed numerically from the total correlation function or vice versa.[44,47,48] The present implementation is instead based on the Fourier-transformed Ornstein-Zernike equation, similar to the approach of Gillan.[49] A grid is introduced in the r and k space. The long-range part of $h_{ij}(r)$ is adjusted, using Newton-Raphson, until the long-range part of $c_{ij}(r)$ coincides sufficiently with the tail-approximation $t_{ij}(r)$. For all systems tested, $h_{\mathrm{MD},ij}(r)$ was used as initial guess and the solution converged within 5–15 iterations.

B. Approximating the long-range direct correlation function

A well-known result for the long-range part of $c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2)$ is[44]

$$c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) \to -\beta u_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2), r_{12} \to \infty, \tag{10}$$

where $u_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2)$ is the pair interaction potential for molecules of type i and j and $\beta \equiv (k_{\mathrm{B}}T)^{-1}$. This also implies that the long-range behavior of the direct correlation function is insensitive to the density. For simulation potentials such as the CHARMM force field,[43] the anisotropic potential at large separation is dominated by dipole-dipole interactions. The potential is written as

$$u_{ij} = u_{ij}^{(\mathrm{LJ})} + u_{ij}^{(\mathrm{dd})}. \tag{11}$$

While Eq. (10) is valid for the molecular direct correlation function, it is not in general for the angle-averaged $c_{ij}(r)$. For example, at low densities, $c_{ij}(r)$ is identical to the angle averaged Mayer f-function[44] which includes a second-order contribution from the dipole-dipole interaction, i.e.,[40]

$$\left\langle \left(u_{ij}^{(\mathrm{dd})}(\boldsymbol{r}_{12}\omega_1\omega_2)\right)^2 \right\rangle_{\omega_1\omega_2} = -\frac{2\mu_i^2\mu_j^2}{3r_{12}^6}. \tag{12}$$

A suitable high-density form is found with the hypernetted chain closure,[50]

$$\begin{aligned} c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) = &-\beta u_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) + h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) \\ &- \log(1 + h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2)). \end{aligned} \tag{13}$$

For a pure fluid composed of rigid, dipolar molecules, Nienhuis and Deutch[51] showed that the asymptotic behavior is

$$h(\boldsymbol{r}_{12}\omega_1\omega_2) \to -\frac{\beta G^2}{\epsilon} u^{(\mathrm{dd})}(\boldsymbol{r}_{12}\omega_1\omega_2), \quad r_{12} \to \infty, \tag{14}$$

where ϵ and G are the dielectric constant and Kirkwood factor of the fluid, respectively. The latter is defined by

$$G \equiv \frac{\langle \boldsymbol{\mu} \cdot \boldsymbol{M} \rangle}{\mu^2}, \tag{15}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{M}$, respectively, denote the dipole moment of a single molecule in the fluid and the total dipole moment of the fluid, and where $\langle \cdot \rangle$ denotes ensemble average. For mixtures, Eq. (14) becomes

$$h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) \to -\frac{\beta G_i G_j}{\epsilon} u_{ij}^{(\mathrm{dd})}(\boldsymbol{r}_{12}\omega_1\omega_2), \quad r_{12} \to \infty, \tag{16}$$

where G_i is a specific Kirkwood factor, defined by

$$G_i \equiv \frac{\langle \boldsymbol{\mu}_i \cdot \boldsymbol{M} \rangle}{\mu_i^2}, \tag{17}$$

where $\boldsymbol{\mu}_i$ denotes the dipole moment of a single molecule of type i. Combining Eq. (16) with the hypernetted chain relation (Eq. (13)) and Taylor-expanding the logarithm of the latter, leads to

$$\begin{aligned} &c_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) \\ &\quad = -\beta u_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2) + \frac{1}{2}(h_{ij}(\boldsymbol{r}_{12}\omega_1\omega_2))^2 + O\left(r_{12}^{-9}\right) \\ &\quad = -\beta u_{ij}^{(\mathrm{LJ})}(\boldsymbol{r}_{12}\omega_1\omega_2) - \beta u_{ij}^{(\mathrm{dd})}(\boldsymbol{r}_{12}\omega_1\omega_2) \\ &\qquad - \frac{\beta^2 G_i^2 G_j^2}{2\epsilon^2}\left(u_{ij}^{(\mathrm{dd})}(\boldsymbol{r}_{12}\omega_1\omega_2)\right)^2 + O\left(r_{12}^{-9}\right). \end{aligned} \tag{18}$$

Only in the last step, the pair potential was assumed to be the sum of Lennard-Jones and dipole-dipole contributions. The final result averages out the angular dependence; to leading order it is

$$c_{ij}(r_{12}) = t_{ij}(r_{12}) + O\left(r_{12}^{-8}\right) \tag{19}$$

with

$$\begin{aligned} t_{ij}(r_{12}) \equiv &-2\beta\left[\sum_{\alpha\in M_i,\beta\in M_j} \epsilon_{\alpha\beta} R_{\mathrm{min},\alpha\beta}^6\right] r_{12}^{-6} \\ &+ \frac{\beta^2 G_i^2 G_j^2 \mu_i^2 \mu_j^2}{3\epsilon^2} r_{12}^{-6}, \end{aligned} \tag{20}$$

where M_i denotes the set of atoms of a molecule of type i and $\epsilon_{\alpha\beta}$, and $R_{\mathrm{min},\alpha\beta}$ denote CHARMM parameters for the Lennard-Jones interaction between atoms of type α and β. The coefficient of the first (Lennard-Jones) term of Eq. (20)

is evaluated directly from the CHARMM parameters. The coefficient of the second (dipole-dipole) term uses G_i, G_j, and ϵ evaluated by Eq. (17) with properties from the simulations and using[52]

$$\epsilon = 1 + \frac{4\pi\beta\rho}{3}\sum_i x_i \mu_i^2 G_i. \tag{21}$$

This requires that the electrostatic forces in the simulations are evaluated using tinfoil boundary conditions. The derivation of Eq. (20) used the hypernetted chain relation (Eq. (13)). The Percus-Yevick relation could also be considered, giving the last term of Eq. (20) as

$$\frac{(2G_iG_j - 1)\beta^2\mu_i^2\mu_j^2}{3\epsilon} r_{12}^{-6}. \tag{22}$$

It appears that the hypernetted chain approximation provides better results. First, integral equation theories for dipolar fluids from the hypernetted chain approximation seem to be more accurate for dense fluids.[53–55] Second, the dipole-dipole contribution to the long-range direct correlation function would be expected to be positive, based on Eq. (12). The hypernetted chain result, Eq. (20), is positive definite while the Percus-Yevick result, Eq. (22), is not. In fact, the Percus-Yevick derived coefficient becomes negative for several of the fluid mixtures we studied.[41] The asymptotic behavior of correlation functions in special situations (e.g., interactions between polar molecules and conducting walls), and possible corrections to the hypernetted chain result, have been discussed in the literature.[56–58]

C. Errors due to finiteness

We discuss here molecular dynamics simulations of the pure Lennard-Jones fluid to illustrate how different types of truncation errors affect the numerical direct correlation function, i.e., the function as computed without employing the extrapolation method of Sec. III A. As we shall see, the extrapolation method corrects satisfactorily for such errors. The simulations used the argon parameters[40] at reduced density (ρ/σ^3) and temperature (k_BT/ϵ) of 0.822 and 1.15, respectively, where ϵ and σ are the Lennard-Jones parameters. The equilibration times were at least 200 ps and the production times were between 800 ps and 8 ns, where systems containing fewer particles were simulated for longer times. The simulations were carried out at constant number of atoms, volume and temperature (NVT-ensemble) using NAMD.[59] The temperature was controlled using a Langevin thermostat with parameters as given in Sec. IV. Radial distribution functions were sampled using a bin width of 0.1 Å. Three simulations comprising, respectively, 2000, 8000, or 20 000 atoms in total were carried out. Lennard-Jones forces were truncated at 12 Å and a switching function was used from 10 Å to smoothly force the potential to zero at 12 Å. One additional simulation comprising 20 000 atoms was carried out with Lennard-Jones forces truncated at 22 Å and using the switching function from 20 Å.

Figure 1(b) compares the direct correlation function evaluated with different upper truncation radii, R_t, for the numerical Hankel transform of the total correlation function

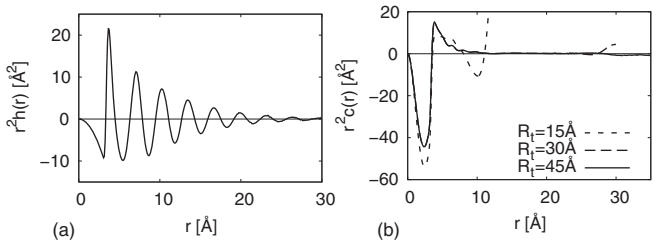

FIG. 1. (a) Total correlation function and (b) numerical direct correlation function of the Lennard-Jones system with $N = 20\,000$. The functions were evaluated by Ornstein-Zernike transformation of the total correlation function truncated at, respectively, 15 Å, 30 Å, and 45 Å (R_t). The line $y = 0$ is marked in both panels to guide the eye.

(Fig. 1(a)). Apparently, $c(r)$ is sensitive to R_t, and with a too small truncation radius (15 Å $\sim 4.4\sigma$), the behavior of $c(r)$ at small r is inaccurate. With $R_t \geq 30$ Å (8.8σ), truncation errors are seen as r approaches R_t with $c(r)$ deviating from the theoretical r^{-6} decay at large r.

The spatial range required for accurate evaluation of $c(r)$ may not be available if the simulated system is too small. In addition, finite-size effects may affect the correlation functions in the accessible range.[31] Size-effects in the total correlation functions (not shown) are very subtle for dense systems with low compressibility. They can result in significant numerical inaccuracies in the reduced bulk modulus and $c(r)$. Figure 2(a) demonstrates this by comparing direct correlation functions obtained using different system sizes. With $N = 2000$, the long range part of $c(r)$ is negative, but shifts upward with increasing N. Figure 2(b) shows that increasing N improves the agreement with the tail approximation (Eq. (20)). With $N = 20\,000$, the agreement is good when $r \geq 6$ Å (1.8σ), which corresponds well to matching distances used in previous applications of the extrapoation method.[40,41] If we replace this $c(r)$ with the approximation for $r \geq 5$ Å, the difference integrated up to 25 Å corresponds to a mere 1% shift in the reduced bulk modulus.

Although the size-effects are difficult to observe in $h(r)$, they are apparent in Fig. 3 showing H, the integral of $h(r)$, as a function of upper truncation limit R_t for systems of different size. With $N = 2000$, the integral seems to diverge and gives a negative value for the reduced bulk modulus. Increasing the system size gradually removes the divergence over the displayed range. The extrapolation method of Sec. III A was

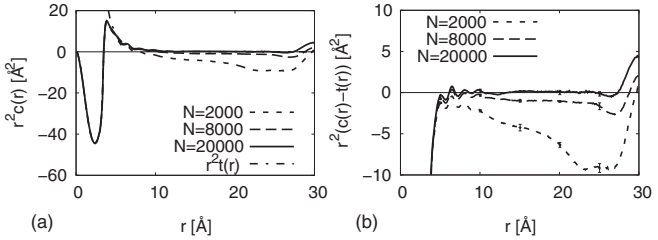

FIG. 2. (a) Numerical direct correlation function of the Lennard-Jones systems with $N = 2000$, 8000, and 20 000, evaluated by Ornstein-Zernike transformation of the total correlation function truncated at 30 Å. (b) Difference between the same direct correlation functions and the long range approximation (Eq. (20)). Standard errors are shown at selected distances and the line $y = 0$ is marked in both panels to guide the eye.

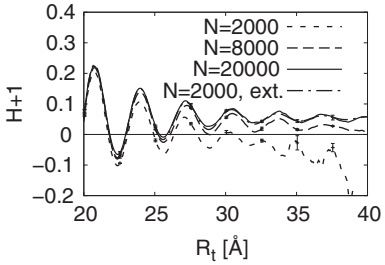

FIG. 3. Reduced bulk modulus, $H + 1$, with H denoting the integral of $h(r)$, of the Lennard-Jones system with $N = 2000$, 8000, and 20 000, evaluated by numerical integration of the total correlation function up to a truncation radius R_t. Corresponding integral for the corrected correlation function for the system with $N = 2000$, obtained by the extension method of Sec. III A with $R = 6.3$ Å, is shown for comparison. Standard errors are shown for selected distances, and the line $y = 0$ is marked to guide the eye.

employed to correct the $h(r)$ obtained with $N = 2000$, using 6.3 Å for the matching distance R of Eq. (9). The integral of the corrected function was in excellent agreement with the integral of $h(r)$ obtained with $N = 20\,000$ (Fig. 3). The extrapolated $h(r)$ integrates to $\rho k_B T \kappa_T = 0.055$, ~2% from the Mecke *et al.*[60] equation of state value of 0.056. This shows that via the extrapolation method, accurate properties can be obtained from simulations of small systems, avoiding the extra computational efforts that larger systems require.

Finally, Figs. 4(a) and 4(b) compare $c(r)$ obtained from simulations using different truncation radii for the Lennard-Jones potential. There is no significant difference between the functions, and they furthermore agree well with the tail approximation (Eq. (20)). For dense fluids, the extrapolation method seems to be insensitive to the use of a truncated Lennard-Jones potential. Also, the tail approximation, which is derived for a potential of unlimited range, seems accurate for a truncated potential.

D. Determining matching distances

The extrapolation method requires appropriate selection of matching distances R_{ij}, as in Eq. (9). The strategy employed here has been described previously,[41] but it is briefly outlined below for completeness.

The parameter R_{ij} is chosen to match $c(r)$ continuously with the tail approximation, $t_{ij}(r)$ (Eq. (20)). The calculations

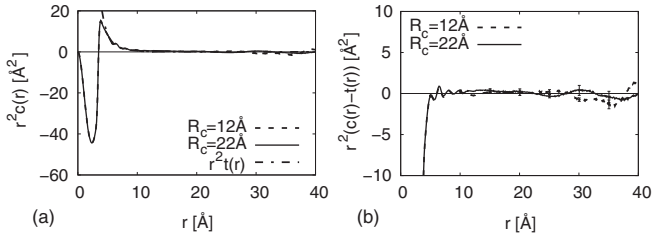

FIG. 4. (a) Numerical direct correlation function of the Lennard-Jones systems with $N = 20\,000$, using potential cutoffs at 12 Å and 22 Å, evaluated by Ornstein-Zernike transformation of the total correlation function truncated at 40 Å. (b) Difference between the same direct correlation functions and the long range approximation (Eq. (20)). Standard errors are shown at selected distances and the line $y = 0$ is marked in both panels to guide the eye.

are therefore first carried out with preliminary parameter values, $R_{ij}^{\dagger}$, set to the largest radius for which $h_{ij}(r)$ is sampled. Preliminary correlation functions $c_{ij}^{\dagger}(r)$ are calculated using the extrapolation method. The final parameters R_{ij} are then chosen at a point beyond the peak of $c_{ij}^{\dagger}(r)$ where it intersects $t_{ij}(r)$. In case $t_{ij}(r)$ does not intersect $c_{ij}^{\dagger}(r)$, R_{ij} is chosen as the value of r after the peak which minimizes the expression,

$$\frac{|c_{ij}^{\dagger}(r) - t_{ij}(r)|}{|t_{ij}(r)|}. \tag{23}$$

The peak of $c_{ij}^{\dagger}(r)$ is approximately independent of $R_{ij}^{\dagger}$, as long as it is large enough for $c_{ij}^{\dagger}(r)$ to be beyond its peak, which ensures that the selection of R_{ij} is robust. For the Lennard-Jones systems shown in Fig. 2, the selected matching distance is 6.1–6.3 Å. For larger r, $t(r)$ agrees very well with the $c(r)$ from the system with $N = 20\,000$, as discussed above.

IV. TESTS FOR MOLECULAR FLUIDS

Molecular models resembling "real" substances most often involve multiple interaction sites. Successful applications to atomic fluids do not necessarily imply that the extension method is accurate for molecular fluids since the treatment of these is less rigorous for two reasons. First, the non-spherical geometry of molecules leads to anisotropic interactions. These interactions might not lead to decoupling of the isotropic and anisotropic correlations as in Eq. (8). In particular, the short-ranged repulsive forces between non-spherical molecules seem to affect the orientation-averaged radial distribution functions to a greater extent than the mostly attractive dipole-dipole interactions, as shown by Monte Carlo simulations of Wang *et al.*[45] Second, the tail approximation for the direct correlation function of Eq. (20) retains only the r^{-6} term in the power series expansion of the angle-averaged pair potential. This term is independent of the molecular geometries as it is determined by the atom types only. Yet, the neglected terms of order r^{-8} and higher do depend on the molecular geometries, becoming more significant as the molecules become more non-spherical.

The ultimate goal of this methodology is to enable efficient simulation-based prediction of thermodynamic properties that are in good agreement with experimental data. This however comprises two tasks. First, a reliable integration procedure must be established. This involves assumptions or approximations of the total correlation functions, and it is thus crucial to validate integration results against previous simulations or alternative simulation-based routes to the same properties. Second, potential models and parameters for the relevant atoms and chemical groups need to be optimized such that experimental properties are reproduced by simulation. For now, we focus on the first step.

Pure atomic fluids and atomic and molecular mixtures have been discussed in previous communications.[40,41] For each test system, the methodology has been applied to evaluate the total correlation function integrals. Depending on the system, the integrals are converted into different

TABLE I. Temperatures, T, pressures, P, and densities, ρ_{MD}, for the simulations of pure molecular fluids. Experimental critical temperatures $T_{c,exp}$ are also listed for each fluid.

	ID	$T_{c,exp}$	T [K]	P [atm]	ρ_{MD} [g/dm^3]
Ethane	EthA	296	380	600	403
	EthB		305	225	399
	EthC		260	100	438
	EthD		180	100	561
	EthE		120	100	636
Butane	ButA	425	500	350	426
	ButB		425	140	424
	ButC		340	100	531
	ButD		260	100	623
	ButE		180	100	704
Hexane	HexA	508	610	300	424
	HexB		508	120	437
	HexC		400	100	561
	HexD		300	100	658
	HexE		200	100	746
Water		647	323	1	1018
2-propanol		509	298	1	781

thermodynamic derivative properties. For instance, for pure fluids isothermal compressibility is evaluated via

$$\rho k_B T \kappa_T = 1 + 4\pi\rho \int_0^\infty r^2 h(r) dr, \tag{24}$$

where ρ, k_B, T, and κ_T, respectively, denote molecular density, Boltzmann constant, temperature, and isothermal compressibility.

We now test the extension method on simulations of pure molecular liquids. The studied liquids were ethane, *n*-butane, *n*-hexane, water, and 2-propanol. Normal alkanes were considered due to their relatively simple structure and because the chain length serves as an order parameter for molecular nonsphericity. Pure water and 2-propanol were considered in order to test the extension method for polar molecular fluids. The calculations are verified by comparing the derivative properties obtained from the integration procedure with the same properties obtained from an alternative "benchmark" method, here compressibility values derived from volume fluctuations.

Simulations of pure ethane, butane, hexane, water, and 2-propanol were carried out in the *NPT* ensemble using the CHARMM force field[43] and the CHARMM-adapted TIP3P model with flexible bonds for water[43] at state conditions summarized in Table I.

The simulations of alkanes and 2-propanol involved 1000 molecules, while the simulation of water included 500. NAMD (Ref. 59) was used for all simulations. The velocity Verlet algorithm with 1 fs time steps was employed to integrate the equations of motion, periodic boundary conditions were employed in the x, y, and z directions, and electrostatic forces were evaluated using the particle mesh Ewald method with a grid spacing smaller than 1 Å. This is consistent with "tinfoil" boundary conditions. Temperature and pressure were, respectively, controlled using the Langevin thermostat algorithm with a damping constant of 5 ps^{-1} and the Langevin piston algorithm with a period of 200 fs and a decay constant of 500 fs. Coordinates were saved every 500 fs. For the water and 2-propanol simulations, Lennard-Jones forces were evaluated using a 12 Å cutoff, a 10 Å switching distance, and a pair list with an outer radius of 14 Å. The alkane simulations used a 15 Å cutoff, a 13 Å switching distance and a pair list outer radius of 17 Å. These were used because some simulations were for state conditions where the fluid is fairly compressible. For such systems, long-range correlations play a more important role than for dense liquids, and the results are more likely to be sensitive to the truncation of Lennard-Jones forces. The systems were equilibrated for at least 200 ps, and the production periods were 8 and 10 ns for the alkane and 2-propanol systems, respectively. For the water systems, the production period was 8 ns. Statistical uncertainties were estimated using the blocking method.[52]

V. RESULTS

The most directly accessible derivative property for pure fluids is the isothermal compressibility. The values of this property obtained from the extension method were compared to those obtained via the fluctuation formula, given by[52]

$$\kappa_T = \frac{\langle V^2 \rangle_{NPT} - \langle V \rangle_{NPT}^2}{k_B T \langle V \rangle_{NPT}}, \tag{25}$$

where V denotes simulation box volume, and $\langle \cdot \rangle_{NPT}$ denotes isothermal-isobaric (*NPT*) ensemble average. The results are listed in Table II. Differences between Eq. (25) and the extension method can be due to many things. In particular, Eq. (24) constitutes an approximate expression where only $h(r)$ appears, instead of the whole (angular dependent) total correlation function. Equation (24) possibly constitutes a source of errors in addition to the various approximations for $c(r)$. It is, based on the present data, difficult to discriminate the relative importance of these approximations.

For the alkane systems, the extension method agreed well with the fluctuation formula with discrepancies in the range of 2–9%. Note that all but one of the calculations resulted in compressibilities that were larger than those obtained from the fluctuation formula. There was no correlation with temperature and density though. As in previous Lennard-Jones and Stockmayer fluids studies,[40] good agreement was seen at supercritical temperatures, as well as for some of the simulations at low temperature, i.e., EthE, ButE, HexD, and HexE. The deviations were only slightly larger than for the pure Lennard-Jones and Stockmayer fluids in the dense region. While this might be due to the non-spherical character of the molecules, the deviations do not seem to increase as the molecules become less spherical. For the water simulation, the result was within 2% which is quite accurate for this strongly anisotropic fluid. The result for 2-propanol deviates 1.4% from the fluctuation value.

The matching distances R, selected for continuity as in Sec. III D, were typically half the upper sampling limit for $h(r)$ or less. This allowed us to verify that the calculated extensions for all systems agreed very well with the

TABLE II. Results for isothermal compressibility for the simulations of pure molecular fluids obtained via the fluctuation formula, Eq. (25), the extension method, Eq. (24) and simple integral truncation (Ref. 5). Matching distances for the extension method, R, are also listed. Standard errors were estimated by the blocking method (Ref. 52).

	$(\rho k_B T \kappa_T)_{\text{fluc}}$	$(\rho k_B T \kappa_T)_{\text{ext}}$	Dev. [%]	$(\rho k_B T \kappa_T)_{\text{trunc}}$	Dev. [%]	R [Å]
EthA	0.183 ± 0.003	0.1934 ± 0.0002	6	0.20 ± 0.03	8	9.5
EthB	0.286 ± 0.004	0.303 ± 0.003	6	0.27 ± 0.06	6	10.4
EthC	0.230 ± 0.003	0.2361 ± 0.0007	3	0.22 ± 0.04	5	10.2
EthD	0.0411 ± 0.0003	0.0446 ± 0.0001	8	0.07 ± 0.14	60	12.3
EthE	0.0143 ± 0.0003	0.0141 ± 0.0001	2	-0.02 ± 0.38	220	11.1
ButA	0.224 ± 0.006	0.2349 ± 0.0005	5	0.219 ± 0.001	2	11.7
ButB	0.37 ± 0.02	0.383 ± 0.003	3	0.3653 ± 0.0004	2	13.3
ButC	0.109 ± 0.001	0.118 ± 0.001	9	0.11 ± 0.01	3	14.5
ButD	0.0357 ± 0.0006	0.0391 ± 0.0002	9	0.04 ± 0.02	14	15.0
ButE	0.0131 ± 0.0003	0.0137 ± 0.0002	5	0.01 ± 0.02	4	14.8
HexA	0.225 ± 0.006	0.238 ± 0.001	6	0.226 ± 0.001	0.4	16.7
HexB	0.330 ± 0.005	0.342 ± 0.006	4	0.308 ± 0.003	7	15.5
HexC	0.083 ± 0.003	0.0889 ± 0.0009	8	0.089 ± 0.004	8	18.2
HexD	0.0274 ± 0.0004	0.0285 ± 0.0003	4	0.03 ± 0.01	14	27.8
HexE	0.0090 ± 0.0001	0.0093 ± 0.0001	3	0.01 ± 0.02	38	27.5
Water	0.0767 ± 0.0007	0.0790 ± 0.0002	2	0.080 ± 0.006	2	9.2
2-propanol	0.0377 ± 0.0006	0.0382 ± 0.0003	1	0.03 ± 0.07	20	13.1

sampled $h(r)$ over the accessible range (data not shown). The agreement was also good for the corresponding running integrals (Figs. 5(a)–5(b)), which are more sensitive to subtle differences in the total correlation functions. Note that the extension method accurately predicted $h(r)$ for a range where significant structure remained present, as apparent in Figs. 5(a)–5(b). The integral over this range contributed significantly to the estimates of isothermal compressibilities which, as discussed above, were satisfactory.

From Table II, especially for hexane, it is apparent that larger molecules require larger matching distances R, as do lower temperatures. For HexE, the R used was only slightly smaller than half the simulation box dimension. For molecules that are larger than hexane and studied at similar temperatures, systems with more than 1000 molecules might need to be used.

We also attempted to obtain the isothermal compressibilities by simply truncating the integrals within the sampled range.[5] A key difficulty is that for systems where $h(r)$ has structure beyond the sampled range, the result is sensitive to the truncation radius R_t. Figure 5(b) demonstrates this; as the upper sampling limit is approached, the running integral still oscillates, and truncation may result in a negative estimate of the isothermal compressibility. Commonly, the sensitivity to the truncation radius is reduced by averaging the running total correlation function integrals over an interval corresponding to one oscillation of $g(r)$.[5] For our dense systems, such averaging reduced the dependence on truncation distance to some extent. The results were sensitive to choice of averaging interval and, in some cases, negative estimates of the compressibility were still obtained. Estimates reported in Table II were computed as the mean of the final maximum and minimum of the running integral within the sampled range (see Fig. 5). Error intervals give half the difference between the final maximum and minimum, thus reflecting how sensitive the result is to truncation radius. For the dense systems, where the results were particularly sensitive to truncation, the method of Theodorou and Suter[30] was used to access $h(r)$ up to $\sqrt{(2)}/2$ times the box length. Higher values of r were not considered due to increasing statistical uncertainties.[30]

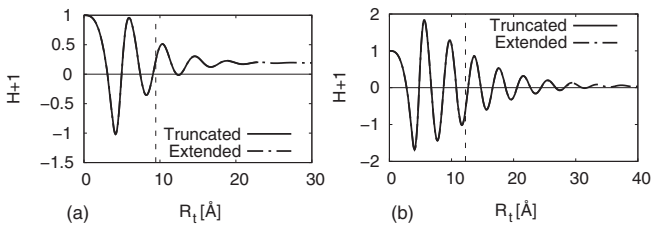

FIG. 5. Running total correlation function integral as function of upper truncation limit (R_t) for (a) Eth1 and (b) Eth4. In both panels, running integrals are shown for the simulation total correlation function (solid) and the extended total correlation function (dashed-dotted), and the line $y = 0$ is shown to guide the eye. The matching distances R for the extension method (see Table II) are marked with vertical dashed lines.

The extension method generally performed better than truncation, although truncation was equally good or better for some of the simulations. For the alkanes, the two methods performed similarly at supercritical temperatures and, in general, for systems with a reduced bulk modulus of at least 0.07. For these systems, the running total correlation function integrals were generally converged within the sampled range (Fig. 5(a)), and integral truncation yielded satisfactory results. For systems with a reduced bulk modulus smaller than 0.05, the truncation results did typically deviate from the fluctuation values with at least 14%. Substantially larger deviations were seen for EthD, EthE, and HexE. For these systems, the uncertainties in the estimates from truncation were comparable or larger than the estimates themselves (Table II), since the running integrals contained significant oscillations of alternating sign beyond the range sampled in simulation (Fig. 5(b)). For the butane family, the extension method is not uniformly

better than truncation. However, Table II seems to suggest that truncation is unexpectedly good for butanes, rather than the extension being appreciably worse for butanes than for other substances. A very good result was obtained for ButE. This is probably coincidental, since the uncertainty is larger than the estimate itself. It was similarly observed for the 2-propanol system, with a reduced bulk modulus of 0.038, that truncation failed due to long-range oscillations, while the extension method performed very well. For water, truncation performed similarly to the extension method, since the integral converged sufficiently within the sampled range.

VI. CONCLUSIONS

We have previously described a computational methodology for extending molecular radial distribution functions obtained from molecular simulation to arbitrarily large spatial separations, so that total correlation function integrals can be reliably obtained by numerical integration. In this paper, this extension method is described in detail and additional testing has been carried out in order to verify that the calculated integrals are accurate for pure molecular fluids. The computed integrals were validated by comparisons with the more rigorous values from density fluctuations. The systems considered for the tests were liquids with $\rho > 2\rho_c$, where ρ_c is the corresponding critical density.

The tests show that the total correlation function integrals are quite accurate under the considered conditions. Good results have been obtained for molecules as nonspherical as hexane and for molecules with anisotropic interactions such as water and 2-propanol. The extension method compares favorably with the simple truncation method,[5] though if the simulated system is sufficiently large, the two methods yield similar results. Thus, the present method is superior when the system is small and the radial distribution functions have significant structure beyond the sampling limit imposed by the simulation box dimension.

ACKNOWLEDGMENTS

The authors gratefully acknowledge access to the Danish Center of Scientific Computing at the University of Southern Denmark, Odense, Denmark. R. Wedberg acknowledges financial support from IP Bioproduction, European Union (EU) Sixth Framework Programme. G. H. Peters acknowledges financial support from the Danish National Research Foundation via a grant to MEMPHYS – Center for Biomembrane Physics.

[1] J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).
[2] J. P. O'Connell, Mol. Phys. **20**, 27 (1971).
[3] J. P. O'Connell, AIChE J. **17**, 658 (1971).
[4] S. Christensen, G. H. Peters, F. Y. Hansen, J. P. O'Connell, and J. Abildskov, Mol. Simul. **33**, 449 (2007).
[5] S. Weerasinghe and P. E. Smith, J. Chem. Phys. **118**, 10663 (2003).
[6] S. Weerasinghe and P. E. Smith, J. Chem. Phys. **119**, 10663 (2003).
[7] S. Weerasinghe and P. E. Smith, J. Phys. Chem. B **107**, 3891 (2003).
[8] S. Weerasinghe and P. E. Smith, J. Chem. Phys. **121**, 2180 (2004).
[9] S. Weerasinghe and P. E. Smith, J. Phys. Chem. B **109**, 15080 (2005).
[10] M. Kang and P. E. Smith, J. Comput. Chem. **27**, 1477 (2006).
[11] N. Bentenitis, N. R. Cox, and P. E. Smith, J. Phys. Chem. B **113**, 12306 (2009).
[12] E. A. Ploetz, N. Bentenitis, and P. E. Smith, Fluid Phase Equilib. **290**, 43 (2010).
[13] S. Christensen, G. H. Peters, F. Y. Hansen, and J. Abildskov, Fluid Phase Equilib. **261** , 185 (2007).
[14] S. Christensen, G. H. Peters, F. Y. Hansen, J. P. O'Connell, and J. Abildskov, Fluid Phase Equilib. **260**, 169 (2007).
[15] H. C. van Ness, J. Chem. Thermodyn. **27**, 113 (1995).
[16] E. Widom, J. Chem. Phys. **39**, 2808 (1963).
[17] J. M. Haile, Fluid Phase Equilib. **26**, 103 (1986).
[18] A. A. Chialvo and J. M. Haile, Fluid Phase Equilib. **37**, 293 (1987).
[19] A. A. Chialvo, J. Chem. Phys. **92**, 673 (1990).
[20] A. Z. Panagiotopoulos, Mol. Phys. **61**, 813 (1987).
[21] A. Z. Panagiotopoulos, Mol. Phys. **62**, 701 (1987).
[22] A. Z. Panagiotopoulos, N. Quirke, M. Stapleton, and D. J. Tildesley, Mol. Phys. **63**, 527 (1988).
[23] K. E. Gubbins and J. P. O'Connell, J. Chem. Phys. **60**, 3449 (1974).
[24] S. W. Brelvi and J. P. O'Connell, AIChE J. **18**, 1239 (1972).
[25] Y. H. Huang and J. P. O'Connell, Fluid Phase Equilib. **37**, 75 (1987).
[26] E. A. Campanella, P. M. Mathias, and J. P. O'Connell, AIChE J. **33**, 2057 (1987).
[27] J. Abildskov, M. D. Ellegaard, and J. P. O'Connell, Fluid Phase Equilib. **286**, 95 (2009).
[28] J. Abildskov, M. D. Ellegaard, and J. P. O'Connell, Fluid Phase Equilib. **295**, 215 (2010).
[29] J. Abildskov, M. D. Ellegaard, and J. P. O'Connell, J. Supercrit. Fluids **55**, 833 (2010).
[30] D. N. Theodorou and U. W. Suter, J. Chem. Phys. **82**, 955 (1985).
[31] J. J. Salacuse, A. R. Denton, and P. A. Egelstaff, Phys. Rev. E **53**, 2382 (1996).
[32] A. Perera and F. Sokolić, J. Chem. Phys. **121**, 11272 (2004).
[33] B. Hess and N. F. A. van der Vegt, Proc. Natl. Acad. Sci. USA **106**, 13296 (2009).
[34] E. Matteoli and G. A. Mansoori, J. Chem. Phys. **103**, 4672 (1995).
[35] R. Wedberg, G. H. Peters, and J. Abildskov, Fluid Phase Equilib. **273**, 1 (2008).
[36] J. W. Nichols, S. G. Moore, and D. R. Wheeler, Phys. Rev. E **80**, 051203 (2009).
[37] F. Lado, Phys. Rev. A **135**, 1013 (1964).
[38] P. Attard, *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximization* (Academic, Amsterdam, 2002), Chap. 9.6, 10.4.2.
[39] S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, D. Bedeaux, and S. Kjelstrup, Chem. Phys. Lett. **504**, 199 (2011).
[40] R. Wedberg, J. P. O'Connell, G. H. Peters, and J. Abildskov, Mol. Simul. **36**, 1243 (2010).
[41] R. Wedberg, J. P. O'Connell, G. H. Peters, and J. Abildskov, Fluid Phase Equilib. **302**, 32 (2011).
[42] L. Verlet, Phys. Rev. **165**, 201 (1968).
[43] A. D. MacKerell, Jr., D. Bashford, M. Bellot, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnik, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wio'rkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. **102**, 3586 (1998).
[44] C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids: Fundamentals* (Oxford University Press, New York, 1984), Vol. 1.
[45] S. S. Wang, C. G. Gray, P. A. Egelstaff, and K. E. Gubbins, Chem. Phys. Lett. **21**, 123 (1973).
[46] J. P. O'Connell, in *Supercritical Fluids: Fundamentals and Applications*, edited by E. Kiran and J. M. H. Levelt Sengers, NATO Science Series E, Vol. 273 (Klüwer Academic Publishers, Dordrecht, Holland, 1994), pp. 191–229.
[47] D. L. Jolly, B. C. Freasier, and R. J. Bearman, Chem. Phys. **15**, 237 (1976).
[48] R. Ramirez, M. Mareschal, and D. Borgis, J. Chem. Phys. **319**, 261 (2005).
[49] M. J. Gillan, Mol. Phys. **38**, 1781 (1979).
[50] D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, New York, 1976).
[51] G. Nienhuis and J. M. Deutch, J. Chem. Phys. **55**, 4213 (1971).
[52] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
[53] S. Murad, K. E. Gubbins, and C. G. Gray, Chem. Phys. **81**, 87 (1983).
[54] P. H. Fries and G. N. Patey, J. Chem. Phys. **82**, 429 (1985).
[55] P. J. Rossky, Annu. Rev. Phys. Chem. **36**, 321 (1985).

[56]G. Stell, *Statistical Mechanics. Part A: Equilibrium Techniques* (Plenum, New York, 1977), p. 47.
[57]P. Attard, J. Chem. Phys. **93**, 7301 (1990).
[58]P. Attard, D. R. Berard, C. P. Usenbach, and G. N. Patey, Phys. Rev. A **44**, 8224 (1991).
[59]J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[60]M. Mecke, A. Müller, J. Winkelmann, J. Vrabec, J. Fischer, R. Span, and W. Wagner, Int. J. Thermophys. **17**, 391 (1996).